



Metadata Interoperability in Public Sector Information

Lina Bountouri¹

Laboratory on Digital Libraries and Electronic Publishing, Department of Archives and Library Sciences, Ionian University, Corfu, Greece

Christos Papatheodorou

Laboratory on Digital Libraries and Electronic Publishing, Department of Archives and Library Sciences, Ionian University, Corfu, Greece and Digital Curation Unit, "Athena" Research Centre, Athens, Greece

Vasilis Soulikias

Local Archive of Limni Evia – State General Archives of Greece, Limni, Evia, Greece

Mathios Stratis

Cataloguing – Department National Library of Greece, Athens, Greece

Abstract

Over recent years, there has been a worldwide growing need for interoperability among the systems that manage and reuse Public Sector Information. This paper explores the documentation needs for Public Sector Information and focuses on metadata interoperability issues. The research work studies a variety of public sector information metadata standards and guidelines internationally accepted and presents two methodologies to obtain interoperability: The first develops an Application Profile, while the second is based on the semantic integration approach and results in the creation of an ontology. The outcomes of the two approaches are compared under the prism of their scope and usage in terms of interoperability during the metadata integration process.

Keywords: Metadata Interoperability; Public Sector Information; Ontology – based integration; Application Profiles; eGovernment

¹ Correspondence to: Lina Bountouri, Laboratory on Digital Libraries and Electronic Publishing, Department of Archives and Library Sciences, Ionian University. 72 Ioanni Theotoki, GR-49100, Corfu, Greece, boudouri@ionio.gr

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

1. Public Sector Information

In recent times the evolution of the information society is influencing the life of citizens in developing and developed countries by providing them with innovative approaches and enhanced technological means to access information. One of the main stakeholders in producing and disseminating huge amounts of information is the public sector. Public Sector Information (PSI) are documents, databases and other information (meteorological information, digital maps, traffic data, etc.) produced, collected, stored, etc. by public sector bodies [1]. This information covers a wide range of domains, such as financial, business and legal. The management of PSI deals mainly with facilitating the Public Administration's (PA) transactions and the access, use and reuse of PSI so as to act as the basis for the provision of services within Public Administration and/or to external users (citizens and enterprises, usually called as "customers") [2].

PSI's effective management is important given that it enables [3]: (a) the reliable documentation of administrative acts, (b) the successful delivery of governments' services by increasing the efficiency of internal processes in PA, facilitating the communication of citizens with the PA and increasing their involvement in the democratic process, (c) the financial progress, since business activities cannot be realized without well-organized and accurate information (quick and easy access to such information facilitate business to make well-informed choices), and (d) the commercial use of PSI, given that creating an improved infrastructure for the exploitation of the PSI facilitates the creation of information products.

Government information is significant for citizens that interrelate with public services on a daily basis and need to be guided and informed promptly in order to easily accomplish their transactions with the public sector [4]. In Greece, for instance, the creation of a government information system by the General Secretariat for Information Systems (Ministry of Economy and Finance) for the internal transactions of the Public Administration and for its external transactions (citizens and enterprises) [5] improved customers' access by giving them the chance to online acquire documents, such as the update for their tax situation, or online depose documents, such as their tax return. In parallel, assisting citizen transactions respects the right to Freedom of Information which defines a legal process by which government information is required to be available to the public. Furthermore, PSI facilitates the decision making process preserving the collective memory of organizations' transactions. Enterprises have to be informed on the latest internal market legal and administrative operations, given that the lack of reliable information could influence smooth and efficient conduct of internal and external trade procedures. In the UK for example, there is a movement towards the improvement of access to PSI, since it supports a large part of the economy and its use and, especially, re-use proves to underpin the decision making processes of customers [6].

One-stop-shop service model has been adopted to fulfil the citizens' needs for completing efficiently their transactions with the public sector. A variety of systems have been developed based on this service model aiming also to satisfy the augmented information needs of government and society. Such systems should follow standards for organizing and disseminating information, standards for the administrative terminology as well as standards/protocols for the communication between systems. Their lack affects and causes simultaneously the heterogeneity and the autonomous development of information systems. Systems may be heterogeneous in levels of [7]:

- syntax: heterogeneities caused by the differences between protocols, encodings and languages used by information sources (i.e. query languages, data formats etc)
- schema: heterogeneities coming from the implementation of different data models, data structures and schemas
- semantic: heterogeneities produced by semantic conflicts arising from the fact that the meaning of the data can be expressed in different ways and with different interpretations
- system: heterogeneities arising from different hardware platforms, operating systems and networking protocols

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

To overcome those issues, it is necessary to achieve *interoperability*. Interoperability is the set of processes ensuring that systems manage their information sources in a way that supports the reuse and exchange of data sources inside and outside a particular system.

One of the main parts of a PSI system that must be interoperable is the metadata schema implemented to cover the documentation needs of PSI resources. Many metadata schemas have been created so as to describe records coming from the public sector in a consistent manner. Most of these schemas are defined and implemented in national level. In detail, until now, there is no particular international metadata schema to document PSI itself. It is worth mentioning the existence of metadata schemas implemented by specific countries for the description of their public records. These schemas are modifications of already existing standards; for example, the Australian Government Locator Service (AGLS) [8], the New Zealand Government Locator Service (NZGLS) [9] and the eGovernment Metadata Standard (e-GMS, used in the UK) [10] are Dublin Core (DC) based [11] and the Global (or Government) Information Locator Service [12] (GILS, used in the USA) is MARC-based.

In order to facilitate interoperability among metadata schemas, various approaches have been developed. Some of the most widely implemented are:

- (1) **Application Profile (AP):** An Application Profile consists of data elements drawn from one or more metadata schemas, combined together by implementers and used for local applications [13]. An AP is based on namespaces, which are used as a stable point of reference to support definition of particular data element sets or vocabularies.
- (2) **Crosswalk:** A Crosswalk is a set of transformations applied to the content of elements in a source metadata standard that results in an appropriately modified content in the analogous elements of a target metadata standard [14]. Specific policies and tools have been developed, with the purpose of converting metadata using crosswalks [15].
- (3) **Ontology based integration:** Ontologies express semantics in a formal manner. They are created to serve as an interoperability mechanism among peoples, institutions and software. In view of the fact that they can conceptualize a domain, they can be used in a metadata interoperability scenario as a mediated schema, as an umbrella of terms and meanings, expressing concepts and the relationships between them [16,17,18,19].

This paper aims to promote interoperability proposing and comparing two approaches. Firstly, we present a methodology for the creation of a Metadata Application Profile for PSI, which is based on the most well-known eGovernment metadata standards. In particular, we propose a model for the description of the PSI as well as the encoding of the administrative public sector constructs that produce information. In our view, a PSI system should be able to encode (a) the information recorded in the public records, (b) the structure and the context of the hierarchy of Public Administration and (c) the structural complexity of its produced archives. To combine these issues, the proposed model suggests (a) the encoding of the records' information via a unified AP, including elements drawn from various eGovernment metadata schemas, (b) the encoding of the collective bodies' and Public Administration's structure and their relationships, deploying the Encoded Archival Context (EAC) metadata schema [20], and (c) the encoding of the Public Records structure using the Encoded Archival Description (EAD) [21] metadata schema.

Secondly, we create an ontology, which is based on the analysis of archival, records and eGovernment information standards in an attempt to act as a reference domain model and mediated schema between systems and metadata schemas documenting PSI. This ontology presents explicitly the main entities that participate to the PSI life-cycle as well as their relations. Finally, we discuss the adequacy of those two methods in terms of interoperability.

The paper is structured as follows: In the next section we propose an AP for PSI, presenting the methodology for its creation, its ability to interlink to sets of records and PA units and the AP itself. To continue, we propose an ontology for documenting and managing PSI and analyze its classes and properties. In the next sections we present a PSI metadata interoperability scenario based on the proposed ontology-based approach. We also present related research efforts and explore the role of each of the proposed approaches (APs and Ontology-based integration). Finally, we present our conclusions and future research efforts.

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

2. An Application Profile for PSI

PSI is produced either via Public Administration's internal procedures or via its communication to external users, such as citizens and business. In both cases, functions are executed and records are produced in order to accomplish the internal or external task/service. As a result, a PSI system must be able to describe both functions and records; hence, it must be based on standards that can encode the structure and the wider context of PA, the complexity and the interlinking of the produced records (and / or archives) and the information included inside the records.

The proposed AP combines elements drawn from the most well-known metadata standards of the records and government information domain, in order to encode PSI. Those metadata standards are: DC, AGLS, NZGLS, e-GMS, GILS and GovML [22] GovML is an XML language able to define structures (or vocabularies) for governmental data and metadata vocabulary so as to support the dissemination of government information produced in the EU; it consists of three vocabularies: (a) generic description data vocabulary for public services, (b) specific description data vocabulary for public services, and (c) data vocabulary for life events and business situations and it also includes an additional GovML vocabulary for metadata elements based mainly on DC. As already mentioned AGLS, NZGLS and e-GMS are DC based standards and GILS is MARC – based. The DC based standards extend DC by adding more specific refinements and new elements, given that DC is general enough to fulfil the documentation and administrative needs of PSI.

The definition of the proposed AP follows the two new recommendations of the Dublin Core Metadata Initiative (DCMI Abstract Model [23] and DCMI Metadata Terms [24]) that place DC on new bases. According to [18] the DCMI Abstract Model includes three related information models: the first one defines the resources to be described, the second one the description itself and the third one the vocabulary (which consists of the DCMI Metadata Terms). DCMI Abstract Model describes the components and constructs that make up an information structure ("DC description set") and how that information structure is to be interpreted. What is more, the Abstract Model adopts the RDF logic and language. Based on the above, the Abstract Model influences the DCMI Metadata Terms recommendation on various levels. Some of the most important points of this influence are: (a) the replacement of the older terminology used for the DCMI grammatical principles with RDF terminology (for example, the use of "*property* or *element*" instead of "*element*", the use of "*property* with *sub-property* of relation" in preference to "*element refinement*", etc.) based on the fact that the described resource is described using one or more *property values*, (b) the addition of formal *domains* and *ranges* (*classes*) to the *properties* with the purpose of defining what *kind* of described resources and *values* are associated with a given *property* and (c) the differentiation between the *Syntax Encoding Schemes* and *Vocabulary Encoding Schemes* (used to be Encoding Schemes) in order to differentiate the *literal values* from *non-literal values* (a literal value can either be indicated explicitly using a syntax encoding scheme or be inferred from the *range of the property*, while the vocabulary encoding scheme indicates the non-literal values).

The objective of this research work is to compose the mentioned standards with the purpose of creating an integrated AP for the management of PSI. This model "is borrowing" properties and sub-properties from the mentioned DC based standards, GILS and GovML in an attempt to express all diverse and current approaches and simultaneously enhance interoperability between standards.

2.1. Methodology

Examining carefully the mentioned standards, it is obvious that they have a great deal of similarities and some differences. In particular, the DC based standards present common points in terms of the properties' use; however, differences are observed in terms of sub-properties' use and of the additional to DC properties that they provide. Further differences exist in the philosophy of the standards, since some of them are more oriented to the management of PSI, while others are oriented to facilitate the citizens in an attempt to access and retrieve PSI.

The proposed methodology develops an AP able to document in completeness the records, taking also into account the PA's structures, which produced them. The main remark from the study of the metadata standards is that semantic overlapping exists between their properties. In detail, when different metadata standards use a set of

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

exactly the same properties then direct mappings between semantic equivalent properties already exist. Further, mappings indicating the relevance of similar properties could be established and finally properties that cannot be mapped – but their semantics are crucial – must be added to the Application Profile.

Based on these findings, the methodology steps for the creation of the AP are (see Figure 1):

- (1) **Comparison of every metadata standard with DC and Addition of extra properties and sub-properties.** As already mentioned most metadata standards included in our scenario are related to DC, hence DC is used as the connecting point of these metadata and the key standard for the creation of the proposed AP. As a result, in the first stage every metadata standard is compared to DC to reveal the semantic similarities and differences between the properties and sub-properties of the metadata schemas. The comparison step allows *the assessment and addition of extra properties and sub-properties* of each involved standard to DC metadata schema.
- (2) **Semantic resolution and harmonization of properties and sub-properties.** The second step involves semantic resolution and harmonization decisions between the properties and sub-properties selected in the first step. The tasks made in this step are (a) to choose the preferred terms of the properties that share equal semantics (b) to select the sub-properties of particular properties and (c) to define the uses of the properties. This stage explores the usages of the different properties and sub-properties and produces an enriched DC-based metadata schema.

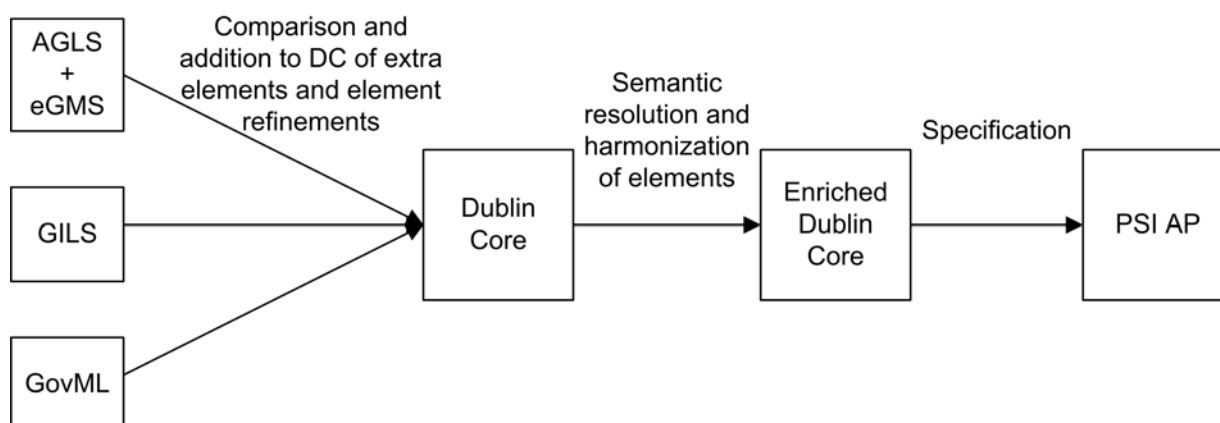


Figure 1. Methodology for the creation of the PSI Application Profile

- (3) **Specification of the PSI Application Profile.** The enriched DC-based schema – due to its large set of terms – needs additional processing in order to avoid repetitions and overlapping. Moreover, new properties are added providing information about the structure of the public records as well as the structure of the PA units that generate the public records.

During the whole methodology the pre-mentioned steps are also taken into account for the vocabulary and syntax encoding schemes that accompany every property participating in the scenario.

The reasons the specific methodological approach is followed are significant. Firstly, the number of the properties, sub-properties, vocabulary and syntax encoding schemes in the AP creation scenario is large enough to easily allow the semantic analysis and the exploration of the similarities and differences of them by executing a single comparison and semantic harmonization step. Secondly, since DC dominates the eGovernment metadata approaches (most of the eGovernment metadata are connected or related to it), it is selected as the main core and stable comparison basis of our effort and thus all the metadata schemas are compared with it.

2.2. Comparison of metadata standards with Dublin Core

DC has a close relationship to the metadata schemas involved: AGLS and e-GMS include almost all its properties (with the exception of one sub-property); GovML includes a part of it, while a crosswalk from GILS to DC already exists. The main findings by the comparison of the metadata standards with DC are:

- Comparison of AGLS and e-GMS with DC. Given that both AGLS and e-GMS include all DC properties (with the exception of the sub-property of the DC Rights property: “Access Rights”), though they have a lot of differences, mainly because of the different documentation and PSI management needs they try to cope with.
- Comparison of GILS with DC. In the specific effort, the comparison process is slightly different than the AGLS – e-GMS – DC comparison, since the one-to-one element mapping is not always the case. For example, the GILS field “Availability”, due to its subfields, is mapped to more than one DC properties. Additionally, GILS – since it is MARC based – does not use any sub-properties and, at the same time, it has many fields that cannot be mapped to DC. These fields were added as extra properties to DC.
- Comparison of GovML with DC. The metadata vocabulary of GovML uses eleven (11) out of sixteen (16) properties of DC; thus, the relationship between the two schemas is obvious. However, the additional three vocabularies of GovML contain properties that can provide valuable information for the description and management of governmental sources. As a result, their addition into the AP is considered essential.

2.3. Semantic resolution and harmonization

The results derived from the semantic resolution and harmonization of AGLS, e-GMS, GILS, GovML and DC are the following:

- Properties representing the same entity but having different names are merged, since they have equal semantics (e.g. “Creator” and “Originator”).
- In case of GILS’ subfields, most of them are regarded as refinements of their parent fields (sub-properties of the properties) given that a subfield refines and specifies the meaning of the field that contains it (e.g. the subfield “Subject Thesaurus” is defined as a refinement of the field “Subject”).
- When two properties are “semantically close” and one of them helps better specifying the other, then the first property is used as a sub-property of the second (e.g. the GILS’ field “Abstract” is used as a sub-property of the property “Description” included in DC, AGLS, and e-GMS).
- With the intention of avoiding overlapping and repetition of properties, some GILS’ fields are excluded since they are already represented by other properties. For instance, the field “Spatial Domain” is excluded because the “Spatial” sub-property of the DC property “Coverage” already wraps the GILS’ field meaning.
- The GovML’s property “Has Translation” is being integrated to the “Relation” property, in view of the fact that it defines the relationship of the described resource to another resource.
- The GovML’s property “Service Code” is being integrated to the “Identifier” property, because it provides identification information.
- The GovML’s property “Law” is being integrated to the “Mandate” property, given that both of them express the law that mandates the creation of a resource or service.
- The GovML’s properties “Delivery Channel”, “Contact Details”, “Public Authority Name”, “Public Authority Department” and “Public Authority Address” are being integrated to “Accessibility” property, since all of them provide information for the accession of the resource or the service.
- Given that GovML does not include any elements’ refinements, there is not any sub-property to be added to the AP by GovML. What is more, none GovML property is used as a sub-property.

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

2.4. *The PSI Application Profile*

The final set created contains fifty-three properties, which is a large set of properties making the model hard to use. Provided that an AP must be a flexible model, additional processing is required. More analytically, in some cases, relevant semantics are represented by more than one property name. As a result, one final name is chosen in order to unify them, while the rest of those properties are either used as sub-properties or totally removed. Many of the non-unified properties may semantically specify other properties. For this reason, it is necessary to be used as sub-properties for them in an attempt to avoid overlapping and repetitions. Numerous properties or sub-properties are used locally from the governmental bodies that created them. Thus, they are removed in order to enhance the interoperability of the model.

Moreover, we propose the usage of (a) EAC for the structural description of the PA units, (b) EAD for the representation of the structure of the PA units' records, and (c) the proposed application profile for the description of the particular records produced by the PA units.

EAC is an XML metadata standard based on International Standard for Archival Authority Records (ISAAR (CPF)) [25]. It provides elements and attributes for the creation of archival authority data describing entities that correlate to the archives' creation, such as personal names, corporate bodies and family names able to depict their relations or structural organization (e.g. a family tree or the structure of an organisation). EAD is the most well known standard for archival description. It is an XML-based descriptive schema, intended to create electronic finding aids, which include the necessary information for the identification, management and interpretation of an archive. EAD and EAC provide linking mechanisms, which permit the interconnection of the structure of a body (e.g. physical, family, corporation, PA units) with their archives. Some of these linking mechanisms are: (a) archive to archive (EAD to EAD), defining the correlation of the archives produced by PA units, (b) archive to authority structure (EAD to EAC), characterizing originator relationship, and (c) authority structure to an authority structure (EAC to EAC), representing complex relationships between PA units, such as superior, inferior and associative units.

The PSI AP is presented in Table 1. The first and the second column represent the properties and sub-properties of the AP respectively; the third column defined the vocabulary and syntax encoding schemes and in the final column properties are placed in four categories depending on their meaning and documentation needs they try to cope with. Three categories of them are defined in [26], which is a clustering of the fifteen core properties of DC. Moreover, the category "Administrative information" is defined for the properties not included in DC in order to represent meanings that are mainly related to PSI management.

Concluding, this information management model correlates the proposed application profile for the PSI description with the information structures (files, sub-files, series, etc.) in which they are included, as well as with the structure of the PA unit, which produces, organises and manages them.

3. **Ontology-based integration of PSI**

Data integration has been an active and challenging research area in the database community for many years. However, nowadays research is moving from data integration to semantic integration in many communities and disciplines, such as information and knowledge management. This movement is being seriously influenced by the new nature and philosophy that the Internet tends to acquire, the Semantic Web [27]. The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to cooperate. The Semantic Web is expected to enable machines to comprehend the semantics of data. The new technologies inspired by the Semantic Web vision, like ontology languages, such as OWL [28] and knowledge representation systems, are continuously offering to the integration field useful tools to exploit and implement in order to semantically integrate data sources.

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

Table 1 The proposed Application Profile

Properties	Sub-properties	Vocabulary and Syntax Encoding schemes	Metadata properties category
Title [DC, AGLS, eGMS, GILS, GovML]	Alternative [DC, AGLS, eGMS]		Content
Creator [DC, AGLS, eGMS, GILS, GovML]		AglsAgent, Government Data Standards Catalogue	Intellectual property
Subject [DC, AGLS, eGMS, GILS, GovML]	Category [eGMS], Person [eGMS], Process Identifier [eGMS], Programme [eGMS], Uncontrolled Term [GILS], Subject Thesaurus [GILS], Subject Terms Controlled [GILS]	LCSh, MeSH, DDC, LCC, UDC, AAT, APAIS, TAGS Category: Government Category List, SIC Keyword: Seamless uk subject Taxonomy, National Curriculum metadata standard, ERIC Person: Government Data Standards Catalogue	Content
Description [DC, AGLS, eGMS, GovML]	Abstract [DC, eGMS, GILS], Table of Contents [DC, eGMS], Purpose [GILS]	URI	Content
Publisher [DC, AGLS, eGMS, GovML]	Place of Publication [GILS]	AglsAgent, Government Data Standards Catalogue	Intellectual property
Contributor [DC, AGLS, eGMS, GILS]		AglsAgent, Government Data Standards Catalogue	Intellectual property
Date [DC, AGLS, eGMS, GovML]	Created [DC, AGLS, eGMS], Modified [DC, AGLS, eGMS], Valid [DC, AGLS, eGMS], Issued [DC, AGLS, eGMS], Date Copyrighted [DC, eGMS], Date Submitted [DC, eGMS], Acquired [eGMS], Available [eGMS], Cut-off Date [eGMS], Closed [eGMS], Date Accepted [eGMS], Declared [eGMS], Next Version Due [eGMS], Updating Frequency [eGMS], Date of Publication [GILS], Date of Last Modification [GILS], Record Review Date [GILS]	DCMI Period, W3C-DTF, ISO8601, ISO19115, Government Data Standards Catalogue	Instantiation
Type [DC, AGLS, eGMS, GovML]	Category [AGLS], Service Type [AGLS], Document Type [AGLS], Automation Level [GovML]	DCMI Type Vocabulary, Agls-document, Agls-Service, e-GMS Encoding Scheme (e-GMSTES)	Instantiation
Format [DC, AGLS, eGMS, GovML]	Extent [DC, AGLS, eGMS], Medium [DC, AGLS, eGMS, GILS]	IMT, AAT, PRONOM	Instantiation
Identifier [DC, AGLS, eGMS, GovML]	Bibliographic Citation [eGMS], Case ID [eGMS], Fileplan ID [eGMS], Control Identifier [GILS], Original Control Identifier [GILS], Service Code [GovML]	URI, ISBN, ISSN, X500, IARN	Instantiation
Source [DC, AGLS, eGMS, GILS]	Record Source [GILS]	URI, ISBN, ISSN	Content
Language [DC, AGLS, eGMS, GovML, GILS]	Language of Record [GILS]	ISO 639-2, RFC 3066	Content
Relation [DC, AGLS, eGMS, GovML]	Is version Of [DC, AGLS, eGMS], Has Version [DC, AGLS, eGMS], Is Replaced By [DC, AGLS, eGMS], Replaces [DC, AGLS, eGMS], Is Required By [DC, AGLS, eGMS], Requires [DC, AGLS, eGMS], Is Part Of [DC, AGLS, eGMS], Has Part [DC, AGLS, eGMS], Is Referenced By [DC, AGLS, eGMS], References [DC, AGLS, eGMS], Is Format Of [DC, AGLS, eGMS], Has Format [DC, AGLS, eGMS], Conforms To [DC, eGMS], Is defined by [eGMS], Provides definition for [eGMS], Reason for redaction [eGMS], Redaction [eGMS], Sequence no [eGMS], Cross Reference Title [GILS], Cross Reference Relationship [GILS], Cross Reference Linkage [GILS], Related Services [GovML]		Content
Coverage [DC, AGLS, eGMS, GovML]	Spacial [DC, AGLS, eGMS, GILS], Temporal [DC, AGLS, eGMS, GILS], Jurisdiction [AGLS], Postcode [AGLS]	DCMI Point, ISO 3166, DCMI Box, TGN, ISO8601, ISO19115, FCO, Government Data Standards Catalogue, AglsJuri, DCMI Period, W3C-DTF	Content

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

Rights [DC, AGLS, eGMS]	Copyright [eGMS], Custodian [eGMS], Descriptor [eGMS], Disclosability to DPA Data Subject [eGMS], DPA Data Subject Access Exemption [eGMS], EIR Disclosability Indicator [eGMS], EIR Exemption [eGMS], FOIA Disclosability Indicator [eGMS], FOIA Exemption [eGMS], FOIA Release Details [eGMS], FOIA Release Date [eGMS], Group Access [eGMS], Individual User Access List [eGMS], Last FOIA Disclosability Review [eGMS], Previous Protective Marking [eGMS], Protective Marking [eGMS], Protective Marking Change Date [eGMS], Protective Marking Expiry Date [eGMS]	Legislation, W3C, Manual of Protective Security Rights	Intellectual property
Audience [DC, AGLS, eGMS, GovML]	Education Level [DC, eGMS], Mediator [DC, eGMS], Addressee [eGMS]	ASCO, ANZSIC, EdNA, Agls-audience, e-GMSAES, IEEE LOM Audience Encoding Scheme	Content
Mandate [AGLS, eGMS]	Act [AGLS], Regulation [AGLS], Authorizing Statute [eGMS], Data Protection Exempt Category [eGMS], Personal Data Acquisition Purpose [eGMS], Law [GovML]		Content
Function [AGLS]	Result [GovML]	AGIFT, Keyword AAA	Administrative Information
Public Authority Type [GovML]	Name of service [GovML]		Administrative Information
Availability [AGLS, GILS]	Available Linkage [GILS], Current Location [eGMS], Home Location [eGMS], Delivery Channel [GovML], Contact Details [GovML], Public Authority Name [GovML], Public Authority Department [GovML], Public Authority Address [GovML], Procedure [GovML], Periodicity [GovML], Time to Deliver [GovML], Cost Info [GovML], Service Hours [GovML], Cost [GovML], Employee Hints [GovML], Citizen Hints [GovML], e-documents [GovML]	AglsAvail	Administrative Information
Accessibility [eGMS]	Access Rights [DC], Originator Dissemination Control [GILS], Security Classification Control [GILS], Eligibility [GovML], Required Documents [GovML], Employee Hints [GovML]	ICRA	Administrative Information
Aggregation [eGMS, AGLS]		IEEE LOM	Content
Digital Signature [eGMS]			Intellectual property
Disposal [eGMS]	Auto Remove Date [eGMS], Disposal Action [eGMS], Disposal Authorised By [eGMS], Disposal Comment [eGMS], Disposal Conditions [eGMS], Disposal Date [eGMS], Date of Last Review [eGMS], Disposal Export Destination [eGMS], Disposal Export Status [eGMS], Disposal Review [eGMS], Disposal Review Details [eGMS], Disposal Reviewer Details [eGMS], Disposal Schedule ID [eGMS], Disposal Time Period [eGMS], Employee Hints [GovML], Citizen Hints [GovML]	U.K. National Archives – Disposal List, W3C-Date Formats	Administrative Information
Preservation [eGMS]	Original Format [eGMS]	PRONOM	Content
Status [eGMS]	State [GovML]	IEEE LOM Status Encoding Scheme	Instantiation Administrative Information
Attention [GovML]			Administrative Information
Methodology [GILS]			Content
Faq List [GovML]			Content

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

Semantic Data Integration is “the process of using a conceptual representation of the data and of their relationships to eliminate possible heterogeneities” [29]. Semantic integration can be considered as a significant part of data integration oriented to solve semantic heterogeneity problems. With the intention of semantically integrating data sources, one has to define the parts of a source “x” that semantically interoperate with the parts of a source “y”.

Ontologies are one of the main infrastructures of the Semantic Web. The most common definition of an ontology is “a formal, explicit specification of a shared conceptualization” [30]. Ontologies can fulfill the lack of shared understanding serving as the basis for (a) modeling an information system communication between different communities of users with different needs, (b) interoperability between systems that have different modeling methods, languages and software platforms, (c) reusability of concepts and relationships among them and reliability by automating the consistency, and d) provision of a common framework across applications for analyzing what entities their data describe [31,32]

In the context of semantic integration, Noy [19] gives a more specialized definition of ontologies saying that an ontology is a formal description of a domain of discourse, intended for sharing among different applications, and expressed in a language that can be used for reasoning.

The last years, there has been much effort to semantically integrate data [29] and metadata [16,17]. Therefore, to fulfil this need, mechanisms that richly express semantics have been used. Ontologies have a vital role in integration efforts - usually acting as the mediated schema - for the reason that they can express semantics in a formal manner [33]. Provided that they can conceptualize a domain, they can be used in a PSI interoperability scenario as a mediated schema, as an umbrella of terms and meanings, expressing concepts and the relationships between them.

For this reason, in the second metadata interoperability approach, we explore the development and use of a core ontology to define the main concepts and relationships for producing and managing PSI. A core ontology can be used to reach an agreement on the entities and their relationships needed for defining a domain or a community of practice. In other words, it can be considered as a formal template to conceptualize the content of a particular domain.

The ontology recommended does not intend to replace metadata schemas that describe public sector or government information, but it defines a conceptual view of PSI which could complement the wider aspect of PSI management, describing the most important concepts and their relations. Thus, the proposed ontology aims to act as (a) a mediated schema in a semantic integration scenario, (b) a reference model for PSI in an attempt to define the main entities that participate to its life-cycle and to explicitly correlate them by specifying a set of relationships (properties), (c) a metadata taxonomy capable to help identifying the requirements for different PSI applications, and (d) a tool to facilitate PSI systems design.

The methodology followed for the ontology development includes three basic steps. To begin with, for the creation of the ontology we decide to follow the *combination development process*. There are three approaches for the creation of an ontology [32]: the *top-down approach*, which starts with the definition of the most general concepts in the domain and subsequent specialization of the concepts, the *bottom-up development process*, which is based on the definition of the most specific classes, the leaves of the hierarchy, with subsequent grouping of these classes into more general concepts, and the *combination development process*, which is a combination of the top-down and bottom-up approaches, by defining the significant concepts first and subsequently generalize and specialize them appropriately. The upper – level concepts we define are *Public Administration*, *Citizen*, *Business*, *Function and Record*, in view of the fact that those classes are the main stakeholders for the production and management of PSI.

Secondly, we deeply study the relevant metadata schemas to explore their “hidden” semantics and the relationships between them. What is more, we do not restrict our study in government information metadata schemas, but we also study records management and archival description standards, such as International Standard on Activities/Functions of Corporate Bodies (ISAF) [34] and General International Standard Archival Description (ISAD (G)) [35]. This action was taken given that PSI is encapsulated in records and archives; ensuring the effective documentation and management of records and archives by applying relevant standards and policies, results in an over time efficient management and exploitation of PSI. Duranti [36] emphasizes in the importance

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

of records for PSI, stating that records play a crucial role in most human activities and they are essential to all of business and social exchanges, they are the basis of the legal system and government functions and accountability depend on them.

Thirdly, we make use of international standards and documentation tools in order to define explicitly the meaning of classes included in the ontology. Some of the standards and tools used are the Suggested Upper Merged Ontology (SUMO) [37], CIDOC CRM [38], WordNet [39], North American Industry Classification System (NAICS) [40], AGLS and GovML.

3.1. The main classes of the ontology

In this section, the scope and content of every class included in the ontology is defined:

1. **Actor:** Person or corporate body that creates, manages, demands, modifies information. The subclasses of Actor class are: (a) *Public Administration*: The Public Administration sector consists of establishments of federal, state, and local government agencies that control and supervise public programs and have executive, legislative, or judicial authority over other institutions within a specified area. These agencies also set strategies, recommend the creation of laws, adjudicate civil and criminal legal cases, and provide for public safety and national defence. Establishments in public sector are typically engaged in the organization and financing of the production of public goods and services [40]. This class can be further analyzed to a taxonomy of classes and subclasses (e.g. Central and Regional administration and then the Central Administration to be analyzed to the classes Ministries and Supervised Public Organizations, while the Regional Administration to the hierarchy Regions, Prefectures, Municipalities, etc.), but this analysis is not of interest for this paper, which focuses on the presentation of the upper level entities of the ontology. (b) *Citizen*: A native or naturalized member of a state or other political community [39]. (c) *Business*: A commercial or industrial enterprise and the people who constitute it [39].
2. **Process:** The class of things that happen and have temporal parts or stages. Examples include extended events like a trial, actions like producing, and biological processes [39]. The subclasses of Process class are: (a) *Function*: According to ISAF, a function is any high-level purpose, responsibility or task, assigned to the accountability agenda of a corporate body by legislation, policy or mandate. Functions are decomposed into a related set of activities. Every function includes activities, which are the tasks performed by a corporate body to accomplish each of its functions. There may be several activities associated with each function. Activities encompass transactions, which in turn produce records. In accordance with ISAF a transaction is the smallest component of an activity and every transaction results in the creation of records, which constitute proof of activities performed to accomplish a function. (b) *Service*: A service exists where a relationship is established between a business function of an organization and the clients' identified needs [8]. A service is divided in two subclasses, since there is a need to represent two related but at the same time distinct (due to the actors that are related to them) notions: *Life-event*: Life events describe situations of human beings that trigger public services [22]. A life-event – in an eGovernment scenario – is an every day life situation of a citizen for which the citizen uses public administration services' in order to confront that situation. Some of the most common life - events are: changing employment status, dealing with crime, having a baby, retiring, dealing with bereavement, moving home, starting / changing school, starting / moving a business. *Business situation*: Business situations describe topics of companies that trigger public services or interactions with public authorities. Examples of business situations are: founding a company, (re-) constructing factory premises etc [22].
3. **Information Object:** This class comprises identifiable immaterial items, such as poems, jokes, data sets, images, texts, multimedia objects, procedural prescriptions, computer program code, algorithm or mathematical formulae that have an objectively recognizable structure and are documented as single units. An Information Object does not depend on a specific physical carrier, which can include human memory, and it can exist on one or more carriers simultaneously [38]. The subclass of Information Object class is: *Record*: In line with [35], a record is information in any form or medium, created or received and maintained by an organization or person in the transaction of business or the conduct of affairs.

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

4. **Law:** A specific warrant which requires the resource to be created or provided [8].
 5. **Policy:** Policy is defined as a plan of action adopted by an individual or social group [37].
- In addition, we define four classes that act as refinements of the pre-mentioned classes, by specializing or better defining their meaning.
6. **Type:** Type class allows additional refinement of classes which require further analysis to represent typological distinctions important to a given community [38].
 7. **Name:** Name class comprises all proper names, words, phrases or codes, either meaningful or not, that are used or can be used to identify a specific instance of some class within a certain context [38].
 8. **Time:** Time class encompasses all proper temporal references (dates, date ranges, etc.)
 9. **Place:** Place class covers extents in space, in particular places of the state that conduct some form of administrative activity. These places could be organised in a taxonomic hierarchy indicating the geographic and geographic division a place belongs.

3.2. The ontology properties

The proposed ontology (Figure 2 and Table 2) is based on three main upper level classes: “Actor”, “Process” and “Information Object”, since those classes encompass the main notions for the production of PSI. More analytically, every “Actor” is linked to the “Process” class via property “*participates in*” and “Process” is linked to “Information Object” via property “*refers to*”, with the purpose of representing that a person or corporate body, which creates, manages, demands, modifies information (“Actor”), is connected to the produced informational entity (“Information Object”) by the use of a specific procedure (“Process”), which promotes the creation, management, demand and modification of information.

The class “Actor” includes three subclasses: “Public Administration”, “Citizen” and “Business”. “Public Administration” is linked to “Record” class via the class-property-class path Public Administration-*carries out*-Function-*generates*-Record. The class “Function” is the connecting point of the information producer (“Public Administration”) and the produced object (“Record”), since a public record is the product of a function executed by the public sector. In archives and records management, the creation of records is connected to functions and not to their producers, given that functions are considered more stable than administrative structures, which are often merged or devolved when restructuring takes place [39]. What is more, employees from different units may cooperate for a specific task and more than one organizational unit may cooperate and produce records for the same activity; in that case, it is difficult to divide the produced records and decide who are the creator and owner. Due to that fact records management and business classification schemes are based on functions. The importance of the “Function” class in PSI information is also proved by the tools and standards published for the explicit use and definition of functions, such as FONZ [41], AGIFT [42] and ISAF.

“Function” is also linked to “Record” by means of the property “*modifies*” to represent that transactions and activities may result in records modification. What is more, it is worth mentioning that “Record” is linked to “Function” via the property “*is evidence of*” since records are proof of activities and transactions. The class “Law” is linked to “Function” through the property “*mandates*” in view of the fact that a legislative or other mandate clarifies the activity producing the records, as it is mentioned in e-GMS and ISAF. Moreover, “Public Administration” is linked to “Law” through the property “*proposes*” in order to show that PA recommends the laws to be created.

Classes “Citizen” and “Business” are both related to PA services (“Life Event” and “Business Situation” respectively) by the use of the relationships “*demand*” and “*requires*”. Life Events and Business Situations are highly important in PSI management, which is proved by the fact that information specialists created particular ontologies in order to define them [42]. Class “Policy” in the specific context comprises strategic activities related to records, such as preservation, appraisal, disposal, rights, accessibility and reproduction policies; hence it is linked to “Record” class through the relationship “*penetrates*”. At the same time, “Public Administration” has an “*adopts*” relationship to “Policy” to define the implementation of policies from the PA.

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

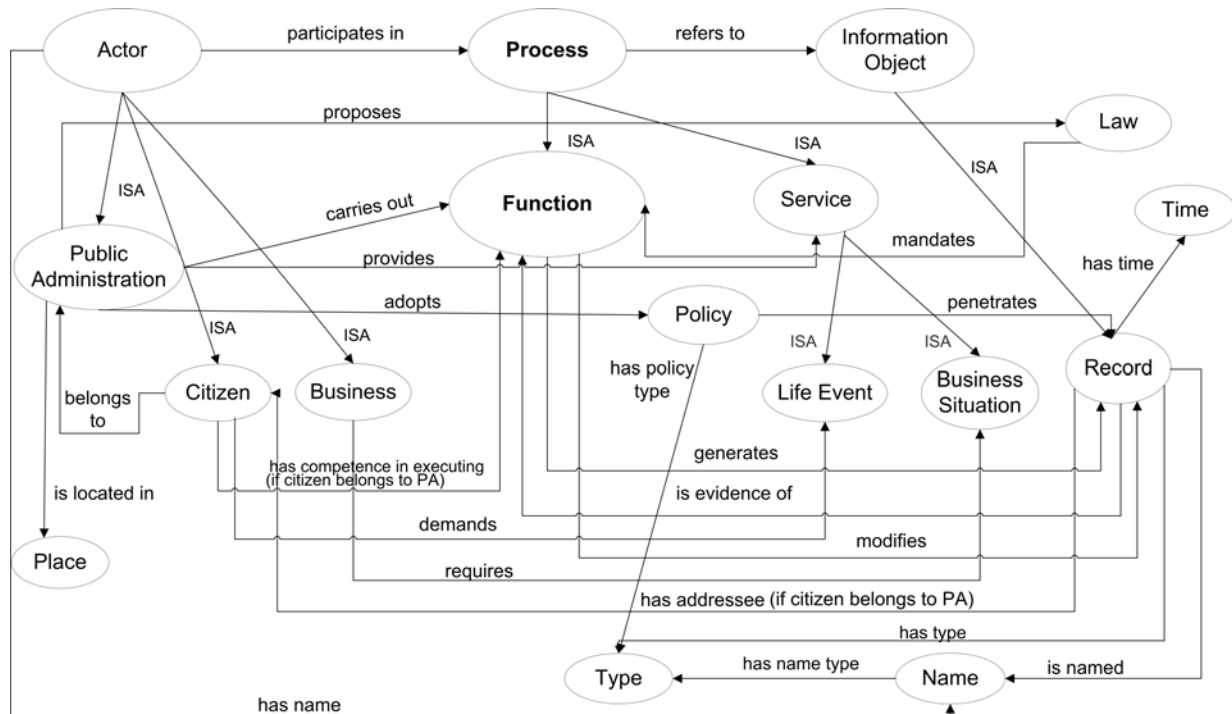


Figure 2. PSI ontology

The additional classes “Name”, “Time” and “Type” are connected to the “Record” class with the properties “is named”, “has time” and “has type” respectively. Furthermore, class “Name” is related to “Type” with a property “has name type” so as to further modify the kind of the name expressed from the specific class. “Public Administration” is related to the class “Place” through the property “is located in” in order to link PA units to the geographical places where they are located. The class “Policy” is also linked to the class “Type” via the “has policy type” property to declare policy specializations, such as accessibility issues, right issues etc.

The class “Citizen” is linked to the class “Function” via the property “has competence in executing”. The specific property is valid only when the property “belongs to” exists between “Citizen” and “Public Administration”, showing that the citizen as an employee of the PA executes functions. What is more, when a citizen is part of the PA, he could also be the addressee of a public record (Record-has addressee-Citizen (if Citizen-belong to-Public Administration)). Finally, “Citizen” is linked to “Public Administration” via the relationship “belongs to”, since in some cases citizens are the employees of Public Administration.

More properties could be added to the proposed ontology model. For example a relation “generates” could be proposed with domain the class “Policy” and range the class “Law” expressing that a policy can generate a law. Nevertheless, such properties are out of the scope of this paper, since they do not add any semantics to PSI documentation and management.

4. Ontology-based metadata interoperability scenario

The proposed ontology defines the complex interrelationships that exist between information objects, actors, functions and other concepts in the PSI field. Given that it is a core ontology, it allows gathering all necessary PSI in a suitable form for further reasoning. In our approach we employ the ontology as a mediator able to semantically integrate diverse PSI sources described with different metadata schemas, enabling the users to

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

retrieve information from them. Thus, we consider PSI ontology as the global schema and we define mappings from the metadata schemas to the ontology and vice versa.

Table 2 Domain and range of the PSI ontology properties

Property	Domain	Range
participates in	Actor	Process
refers to	Process	Information Object
carries out	Public Administration	Function
demands	Citizen	Life Event
requires	Business	Business situation
has competence in executing	Citizen (if Citizen belongs to Public Administration)	Function
generates	Function	Record
modifies	Function	Record
is evidence of	Record	Function
has addressee	Record	Citizen (if Citizen belongs to Public Administration)
proposes	Public Administration	Law
mandates	Law	Function
penetrates	Policy	Record
adopts	Public Administration	Policy
is named	Record	Name
has time	Record	Time
has policy type	Policy	Type
has type	Record	Type
has name type	Name	Type
is located in	Public Administration	Place
belongs to	Citizen	Public Administration
has role	Citizen	Public Administration

Users can pose their queries to the mediator, which transforms them to the local sources' query languages, making use of the defined mappings, and propagates them for execution to the local sources. The results from every source in demand are promoted to the mediator, which translates them (using again the appropriate mappings) and are returned to the user. Note that the queries to the ontology as well as the metadata sources might be written in a query language such as SPARQL and XQUERY depending on the encoding syntax of every metadata schema. For example, suppose a user wish to find metadata records describing documents created by a PA unit named "Central Administrative Unit of Educational Media". Suppose that the user poses its (appropriately formed) query to the PSI mediator ontology. Then, the query is transformed in terms of the elements of the local sources' metadata schemas, using the appropriately defined mappings and, then, sent and executed to the local data sources. A repository hosting e-GMS records will return the records for which the element "Creator" equals to the value "Central Administrative Unit of Educational Media". The e-GMS records that match the query will be propagated to the mediator and transformed, using the mapping rules from e-GMS to PSI ontology, into an equivalent result set in terms of the ontology. The same may occur to a number of sources following different than e-GMS schemas, without the need to map them to each other.

As an indicative example how the proposed ontology integrates PSI resources from different metadata schemas, we define mappings of e-GMS to the ontology, using a path-oriented methodology (see Figure 3). The specific methodology has been also applied in [16,17] for the definition of semantic mappings in the cultural heritage field. In Figure 3, boxes represent the elements of e-GMS (upper part of the box) as well as the classes of the proposed ontology (lower part of the box). Arrows represent the ontology's properties.

In most cases, a mapping from a source schema to a target schema transforms each instance of the source schema into a valid instance of the target schema [43]. Our approach maps *metadata paths* to semantic equal *ontology paths*. A metadata path is defined as a sequence of elements, sub-elements (or element refinements), encoding and vocabulary schemes, starting from the metadata schema root element separated by the slash symbol

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

(/). For the above mentioned example, the path “/eGMS/Creator” is a part of the e-GMS metadata denoting the name of the creator of a public record.

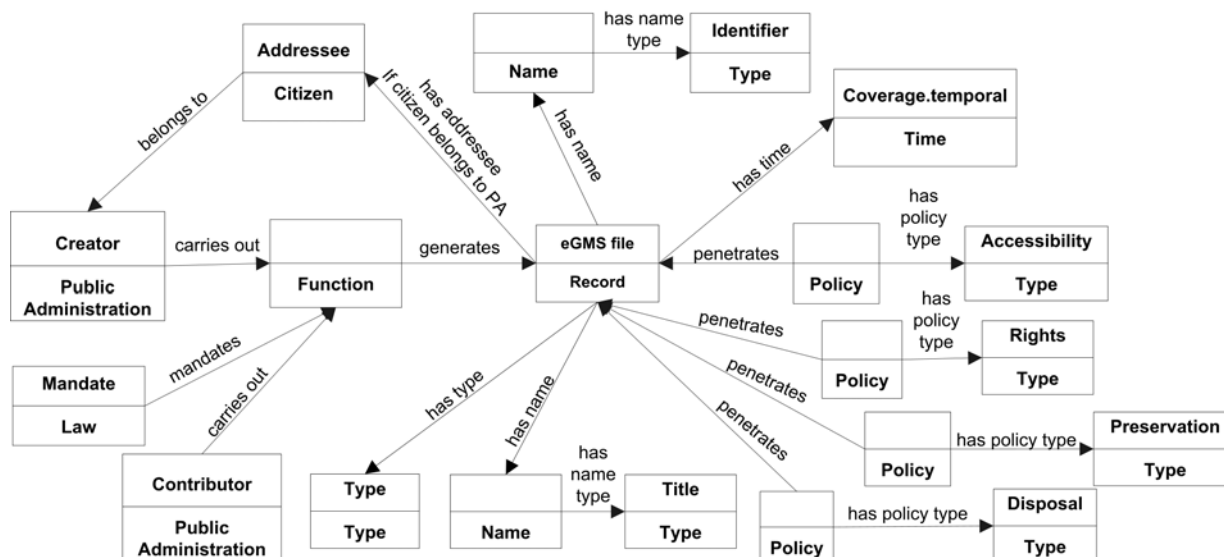


Figure 3. Mapping e-GMS to PSI ontology

On the other hand an ontology path is characterized as a sequence of the form *class-property-class*, such that the classes associated by a property correspond to the property's domain and range. The corresponding ontology path for the path “/eGMS/Creator” of our example is: “Public Administration-carries out-Function-generates-Record” denoting that a creator (Public Administration) carries out a set of tasks (Function), which generate an informational entity (Record). According to Figure 3 the Identifier of a particular record (metadata path: “/eGMS/Identifier”) is mapped to the ontology path: “Record-has name-Name-has name type-Type”. Likewise, the Mandate that recommended the creation of a resource (metadata path: “/eGMS/Mandate”) is mapped to the ontology path: “Law-mandates-Function-generates-Record”. Similarly, the public sector unit that contributes to the resource (metadata path: “/eGMS/Contributor”) corresponds to the ontology path: “Public Administration-carries out-Function-generates-Record”.

As a final point, this integration approach maps the metadata elements to the ontology classes and properties and in the same time it reveals their semantic co-relations. Furthermore, the mapping process could contribute to the further specialization of the ontology classes and properties, so that to be developed a reference model for the PSI.

5. Related Work and Discussion

Since the exploitation and reuse of PSI is becoming really important in Europe [2], the European Committee for Standardization (CEN) proposed a metadata application profile for encoding government information in Europe, based on DC, named CWA 14860: 2003 – eGovernment Application Profile (AP) [44].

The methodology to define the specific AP is the following: (a) Review of the existing metadata standards and initiatives. The first version of the DCAP is based on the metadata standards of United Kingdom, Denmark, Finland, Iceland and Ireland and other relevant standards and initiatives were also taken into account, such as USA, Canada, Australia, and New Zealand. Still none element from the outside of Europe metadata standards is used. (b) Identification of additional - to those specified in DC - metadata terms. Comparison tables were created with the additional metadata properties and sub-properties used in the investigated metadata schemas and not included in DC. To allow comparisons, a table was also created with the definitions of all metadata properties used in DC and member states standards (total number 34 properties). (c) Harmonization. The uses of the

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

metadata properties were defined. (d) Definition of the metadata terms (version 1.0). Compared to DC, this AP includes seven additional properties: accessibility, disposal, location, preservation, status, mandate and metametadata. From the initial eighteen properties in use in Member states, one is currently eliminated, seven are retained as properties while the remaining ten are either included as sub-properties or mapped to those retained. (e) Mapping to existing standards. They created a table with properties that do not obviously have mappings to other standards, i.e. properties with a name and definition that is not identical to the name and definition in the proposed AP.

Comparing the CWA 14860: 2003 – eGovernment Application Profile (AP) with our proposed AP (see Table 3), it should be remarked that the CWA 14860 does not include: properties and sub-properties that (a) support the promotion of eGovernment and the provision of services to users, for instance Service Code, Name of service, Service Hours, Public Authority Type, FaqList, Attention and Related Services (GovML) and (b) document the use of PSI inside and outside the public sector, for example Employee Hints, Citizen Hints (GovML), Language of Record (GILS). Additional missing concepts are those related to the means used to obtain the resource, or contact information for obtaining the resource (Availability (AGLS, GILS)), the authentication information used for the verification of resources in transactions (Digital Signature (eGMS)) and the tools, techniques or methodology implemented for the creation of the PSI resource (Methodology (GILS)).

Besides, basic notions of PSI management are absent, such as the notion of life-events and business situations that correspond to the services offered from public sector to citizens and business. Another significant notion essential for the documentation of basic procedures of the public sector, which is not included in CWA 14860, is the property “Function”. Public Records are created and disseminated into the Public Administration and reflect its procedures and structure. In other words the information produced by G2C, G2B and G2G transactions, is highly dependent on Public Administration structure and activities. Our Application Profile is based on this approach, which we believe to be citizen-centric. Therefore the introduced property Function describes the activities and transactions that produce the records, hence PSI, and it is considered to be the linking point of records and the units that produce them. The existent standards view the Public Records from an administrative point of view (i.e. an information management approach) and do not follow the “functional approach”. Though several metadata standards (such as AGLS, NZGLS) follow the functional approach, the existent Application Profiles, such as CWA 14860, miss this information and cannot integrate it.

It is also really important to mention that the comparison and semantic resolution and harmonization step of the proposed methodology for the development of the PSI AP, allowed us to better define the semantics of properties and sub-properties and avoid repetitions. For example, in the CWA 14860 Custodian is used as a sub-property of both the Creator and Rights properties, while its precise semantic declares the user or role identifier with local management powers over the resource, e.g. assignment and maintenance of access control markings. In our AP Custodian is used as a sub-property of the property Rights since its meaning is mostly oriented to access, use and control issues.

In the same vein, the Dublin Core Metadata Initiative has also proposed an AP for eGovernment metadata [45], which integrates the metadata from several countries such as UK, Australia, Finland, etc. Nonetheless, the work of the Dublin Core Working Group (DC-Gov) is currently under way and at the present it is not known when this process will be completed and if the AP will be accepted by the Usage Board.

An additional effort related to our ontology-based approach is presented in [46]. A conceptual metadata schema model for record management is defined, based on record management guidelines of ISO 15489 [47],[48] and ISO 23081 [49]. This schema’s target is to maintain international compatibility and standard management procedures. It is a record-centred model consisted of three basic elements: Records, Business, and Mandate. However, this model is orientated to records management and due to that fact, even if it can deal with parts of PSI, it is not related to basic notions of eGovernment, such as services offered from the PA (life events and business situations).

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

Table 3 Comparison of the CWA 14860: 2003 – eGovernment Application Profile to the proposed PSI Application Profile

CWA 14860 Properties	CWA 14860 Sub-properties	PSI AP Properties	PSI AP Sub-properties
dc: Title	dc: alternative	Title [DC, AGLS, eGMS, GILS, GovML]	Alternative [DC, AGLS, eGMS]
dc: Creator	UK: Owner, contact, custodian FI: PersonalName, corporateName	Creator [DC, AGLS, eGMS, GILS, GovML]	
dc: Subject	UK: Category, keyword, processIdentifier, programme, project	Subject [DC, AGLS, eGMS, GILS, GovML]	Category [eGMS], Person [eGMS], Process Identifier [eGMS], Programme [eGMS], Uncontrolled Term [GILS], Subject Thesaurus [GILS], Subject Terms Controlled [GILS]
dc: Description	dc: TableOfContents, abstract	Description [DC, AGLS, eGMS, GovML]	Abstract [DC, eGMS, GILS], Table of Contents [DC, eGMS], Purpose [GILS]
dc: Publisher		Publisher [DC, AGLS, eGMS, GovML]	Place of Publication [GILS]
dc: Contributor	FI: PersonalName, corporateName IC: Requirements	Contributor [DC, AGLS, eGMS, GILS]	
dc: Date	dc: Created, valid, available, issued, modified, dateAccepted, dateCopyrighted, dateSubmitted UK: Acquired, CutOffDate, closed, declared, nextVersionDue, updatingFrequency DK: DateSend, DateAction FI: Accepted, dataGathered, retentionPeriod	Date [DC, AGLS, eGMS, GovML]	Created [DC, AGLS, eGMS], Modified [DC, AGLS, eGMS], Valid [DC, AGLS, eGMS], Issued [DC, AGLS, eGMS], Date Copyrighted [DC, eGMS], Date Submitted [DC, eGMS], Acquired [eGMS], Available [eGMS], Cut-off Date [eGMS], Closed [eGMS] Date Accepted [eGMS], Declared [eGMS], Next Version Due [eGMS], Updating Frequency [eGMS], Date of Publication [GILS], Date of Last Modification [GILS], Record Review Date [GILS]
dc: Type	UK: FolderType CEN: Aggregation, DocumentType	Type [DC, AGLS, eGMS, GovML]	Category [AGLS], Service Type [AGLS], Document Type [AGLS], Automation Level [GovML]
dc: Format	dc: Extent, Medium, IMT	Format [DC, AGLS, eGMS, GovML]	Extent [DC, AGLS, eGMS], Medium [DC, AGLS, eGMS, GILS]
dc: Identifier	dc: bibliographicCitation UK: FileplanID, systemID	Identifier [DC, AGLS, eGMS, GovML]	Bibliographic Citation [eGMS], Case ID [eGMS], Fileplan ID [eGMS], Control Identifier [GILS], Original Control Identifier [GILS], Service Code [GovML]
dc: Source		Source [DC, AGLS, eGMS, GILS]	Record Source [GILS]
dc: Language		Language [DC, AGLS, eGMS, GovML, GILS]	Language of Record [GILS]
dc: Relation	dc: IsVersionOf, hasVersion, isReplacedBy, replaces, isRequiredBy, requires, isPartOf, hasPart, isReferencedBy, references, isFormatOf, hasFormat, conformsTo UK: IsDefinedBy, providesDefinitionOf, reasonForRedaction, redaction, sequenceNo	Relation [DC, AGLS, eGMS, GovML]	Is version Of [DC, AGLS, eGMS], Has Version [DC, AGLS, eGMS], Is Replaced By [DC, AGLS, eGMS], Replaces [DC, AGLS, eGMS], Is Required By [DC, AGLS, eGMS], Requires [DC, AGLS, eGMS], Is Part Of [DC, AGLS, eGMS], Has Part [DC, AGLS, eGMS], Is Referenced By [DC, AGLS, eGMS], References [DC, AGLS, eGMS], Is Format Of [DC, AGLS, eGMS], Has Format [DC, AGLS, eGMS], Conforms To [DC, eGMS], Is defined by [eGMS], Provides definition for [eGMS], Reason for redaction [eGMS], Redaction [eGMS], Sequence no [eGMS], Cross Reference Title [GILS], Cross Reference Relationship [GILS], Cross Reference Linkage [GILS], Related Services [GovML]
dc: Coverage	dc: Spatial, temporal FI: jurisdiction	Coverage [DC, AGLS, eGMS, GovML]	Spatial [DC, AGLS, eGMS, GILS], Temporal [DC, AGLS, eGMS, GILS], Jurisdiction [AGLS], Postcode [AGLS]
dc: Rights	dc: accessRights IC: Price DK: RightsSecurityClassification UK: Copyright, custodian, descriptor, disclosabilityToDPADDataSubject, DPADDataSubjectAccessExemption, EIRDisclosabilityIndicator, EIRExemption, FOIDisclosabilityIndicator, FOIExemption,	Rights [DC, AGLS, eGMS]	Copyright [eGMS], Custodian [eGMS], Descriptor [eGMS], Disclosability to DPA Data Subject [eGMS], DPA Data Subject Access Exemption [eGMS], EIR Disclosability Indicator [eGMS], EIR Exemption [eGMS], FOIA Disclosability Indicator [eGMS], FOIA Exemption [eGMS], FOIA Release Details [eGMS], FOIA Release Date [eGMS], Group Access [eGMS], Individual User Access List [eGMS], Last FOIA Disclosability Review [eGMS], Previous Protective Marking [eGMS], Protective Marking [eGMS], Protective Marking Change Date [eGMS], Protective

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

	FOIReleaseDetails, FOIReleaseDate, groupAccess, individualUserAccessList, lastFOIDisclosabilityReview, previousProtectiveMarking, protectiveMarking, protectiveMarkingChangeDate, protectiveMarkingExpiryDate CEN: Publicity		Marking Expiry Date [eGMS]
dc: Audience	dc: Mediator, educationLevel CEN: Receiver	Audience [DC, AGLS, eGMS, GovML]	Education Level [DC, eGMS], Mediator [DC, eGMS], Addressee [eGMS]
UK: Accessibility		Accessibility [eGMS]	Access Rights [DC], Originator Dissemination Control [GILS], Security Classification Control [GILS], Eligibility [GovML], Required Documents [GovML], Employee Hints [GovML]
UK: Disposal	UK: AutoRemoveDate, disposalAction, disposalAuthorisedBy, disposalComment, disposalConditions, disposalDate, dateOfLastReview, disposalExportDestination, disposalExportStatus, disposalReview, disposalReviewDetails, disposalReviewerDetails, disposalScheduleID, disposalTimePeriod	Disposal [eGMS]	Auto Remove Date [eGMS], Disposal Action [eGMS], Disposal Authorised By [eGMS], Disposal Comment [eGMS], Disposal Conditions [eGMS], Disposal Date [eGMS], Date of Last Review [eGMS], Disposal Export Destination [eGMS], Disposal Export Status [eGMS], Disposal Review [eGMS], Disposal Review Details [eGMS], Disposal Reviewer Details [eGMS], Disposal Schedule ID [eGMS], Disposal Time Period [eGMS], Employee Hints [GovML], Citizen Hints [GovML]
UK: Location	UK: CurrentLocation, homeLocation		
UK: Preservation	UK: originalFormat	Preservation [eGMS]	Original Format [eGMS]
UK: Status	CEN: Version	Status [eGMS]	State [GovML]
UK: Mandate	UK: AuthorisingStatute, dataProtectionExemptCategory, personalDataAcquisitionPurpose CEN: Mandator	Mandate [AGLS, eGMS]	Act [AGLS], Regulation [AGLS], Authorizing Statute [eGMS], Data Protection Exempt Category [eGMS], Personal Data Acquisition Purpose [eGMS], Law [GovML]
CEN: MetaMetadata	CEN: Creator, dateCreated, dateModified		
		Methodology [GILS]	
		Faq List [GovML]	
		Digital Signature [eGMS]	
		Function [AGLS]	Result [GovML]
		Public Authority Type [GovML]	Name of service [GovML]
		Availability [AGLS, GILS]	Available Linkage [GILS], Current Location [eGMS], Home Location [eGMS], Delivery Channel [GovML], Contact Details [GovML], Public Authority Name [GovML], Public Authority Department [GovML], Public Authority Address [GovML], Procedure [GovML], Periodicity [GovML], Time to Deliver [GovML], Cost Info [GovML], Service Hours [GovML], Cost [GovML], Employee Hints [GovML], Citizen Hints [GovML], e-documents [GovML]
		Attention [GovML]	

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

Moreover, as far as the ontology approach is concerned, it is important to notice that there is no core ontology for PSI management, even though from the literature review ontologies for related disciplines have been found. In [50], a Public Administration ontology is proposed representing the classic views of PA: (a) legal view (legal rules relating to PA), (b) organizational view (organizational rules, business processes, institutions etc), (c) business view (business rules), (d) IT view (data in PA and their management, web portals etc), and e) end-user view (services for citizens and business). Nevertheless, this ontology is adapted as a part mechanism inside a life-event portal, it includes very broad concepts and, since it is not based on records management and archival policies, it is inadequate to be used as a mediated technology between records, archives and PSI documentation.

In OntoGov project [51], various ontologies are proposed (i.e. Lifecycle ontology, Legal ontology, Process ontology, Profile Ontology) and used for modelling eGovernment services, aiming mainly to interpret and adapt at the conceptual level changes of the eGovernment system that are result of the continuously changing legal environment. The goal of the specific project is to develop and support the back – office procedures by creating a system that can frequently update itself in accordance with the changes in the domain. This approach utilizes CWA 14860 to model information; nonetheless it does not cope with the metadata interoperability issue. Finally the incorporation of existing upper level ontologies such as SUMO [37] and DOLCE [52] in the PSI domain was considered out of the scope, given that they are oriented to define knowledge in a wide context and to act as tools from which core ontologies could be extracted.

Our methodology for the definition of the PSI ontology origins by the study of the existent metadata standards, allowed us to mix and inter-cross concepts from a diversity of metadata schemas and, consequently, produce a tool able to cope with various PSI documentation needs. The PSI ontology defines an upper level conceptualization of semantics and identifies the logic onto which the resources are created. Therefore it could be used as a formal domain model on which a Description Set Profile [53] of an Application Profile should be based. An ontology is a domain model which defines the entities that the resources belong as instances and their relationships, while the Description Set Profile of an Application profile declares the metadata vocabularies (in terms of properties) and their structural and syntactic constraints.

Core ontologies can be considered as an important building block for integration architectures, into which metadata originating from diverse sources can be semantically mapped and integrated [54]. Some of the most known core ontologies are: (a) ABC Ontology [55], which has been developed in order to provide a common conceptual model to facilitate the interoperability between metadata ontologies from different domains, and (b) CIDOC CRM, which provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage. Related work and use of those two core ontologies as an interoperability mechanism are presented in [56],[17]. ABC Model in particular is used to formally define shared entities and relationships that exist in multiple metadata vocabularies, logically describe those entities and their inter-relationships, and it also can be used as a framework for extending shared semantics to domain- and application-specific metadata vocabularies.

Based on the above APs and ontologies are considered as a vital part of metadata interoperability scenarios. Nonetheless, APs and ontologies differ in terms of:

1. **Scope:** APs are object oriented and focus on covering the search and retrieval needs of users looking for PSI records by providing attributes to the record so as to facilitate its identification. On the other hand, ontologies target to specify the semantic notions that define a domain (e.g. cultural heritage, multimedia, legal etc.) and their interrelationships providing a wider yet explicitly specified aspect, mostly oriented to cover information integration needs. For example the basic notions of *function*, *PA* and *services* surround the notion of a *record* as secondary parts, while in the ontology they have the same semantic equivalence and role, providing a well-defined view of the PA and the life-cycle of the record. The APs provide the notion of record with extra descriptive properties and subproperties, whilst ontologies organize these properties to semantic classes revealing also explicitly their interrelations. Hence, through an ontology specific rules over the classes and properties can be declared, creating a rich semantic network. An indicative example is the declaration that a record was either produced or modified by a PA function (i.e. the relationships “*generates*”, “*modifies*”, “*is evidence of*”, etc. between the classes “*Function*” and “*Record*”), which is mandated by a specific law (i.e. the relationship “*mandates*” between the classes “*Law*” and “*Function*”). In an AP record the existence of mandate and function (Mandate and Function properties in the PSI AP) are defined without declaring the

relationship which may have not only with the described resource (the record) but also between them or with other basic semantic PSI notions, such as “Public Administration”. What is more, in the PSI AP the role of the PA is not obviously indicated. In the ontology actions executed by the PA are defined, for instance a “Law” is proposed by the “Public Administration” and this “Law” “*mandates*” the “Function” which will finally affect the creation, modification etc of the record. Also, in the PSI AP the proposed properties and sub-properties related to persons (for example `dc.creator.contact` in the CWA 14860: 2003 AP) are defined without any explicit and machine-readable intermediate connection between them and the wider category they belong, such as the Public Administration.

Another significant issue is that ontologies view the information sources as instances of particular classes. In short, they reveal the meaning of the metadata (information), being more abstract constructs than metadata schemas, hence revealing the relationships not only between classes but also between various metadata schemas used for encoding records and archives. Therefore, various metadata semantics are included under the definition of ontology’s classes, such as ISAF semantics under the class “Function”, EAC semantics which under the class “Actor” and eGMS under the class “Record”.

Concluding, ontologies cannot replace the use of metadata schemas and APs and it is worth mentioning that metadata have a completely different scope and function in comparison with ontologies. They are used to describe, identify, and facilitate the access, usage and management of information. Ontologies intend to conceptualize a field of interest by defining the classes and their properties as well as the classes’ relationships.

2. **Interoperability:** As we have already mentioned, in the Semantic Integration field conceptual representations of the metadata and their relationships are used to cope with heterogeneity issues. Ontologies can be easily used in a data integration system with the aim of describing and defining the semantics of the data sources and making their content explicit [18],[57]. Their nature allows (a) the sophisticated, extended and rich expression of meanings, (b) the modeling of mappings between meanings, (c) the ability of reasoning, (d) a higher degree of abstraction, as the model is separated from the data storage and (e) a query model since it corresponds to the users’ view of the domain. In our proposed integration scenario, the core ontology is used as a semantic mediator and query model; its role is important in the context in which user needs may not be referred to only one data source but instead their satisfaction depends on combinations of multiple heterogeneous data sources with different representations of a common domain [58],[59]. The proposed ontology is – as already mentioned – a rich descriptive semantic network able to integrate metadata schemas, since metadata schemas can act as instances of its classes.

On the other hand, APs promote interoperability since their description core is derived from already used and defined metadata schemas, such as Dublin Core, and in terms of using already known metadata terms from other schemas. In detail, an AP is based on one single schema but tailored to different user communities, such as the DC-Library Application Profile that defines the use of the DC metadata schema in library-related applications and projects [60]. The AP’s purpose is to adapt or combine existing schemas into a package that is tailored to the functional requirements of a particular application, while retaining interoperability with the original base schema or schemas. Part of such an adaptation may include the elaboration of local metadata elements that have importance in a given community or organization, but which are not expected to be usable in a wider documentation context [61].

A significant difference between APs and ontologies is that possible changes in the description needs usually affect the definition and the metadata terms of an AP, since those changes must be adopted and represented by an AP. For instance, a change in the DC base metadata schema must be “pictured” in the AP in order to continue promoting interoperability to other DC based standards and satisfy the description needs. However, core ontologies define and represent the abstractness of a specific domain; hence they are usually free of syntactic structures and mainly of representing the changing documentation needs and adopting any modifications. For this reason, ontologies considered to be more stable constructs over time.

What is more ontology-based integration is an adequate metadata integration solution with reference to other popular metadata interoperability methods, such as Crosswalks. In Crosswalks a polynomial increase of

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

mappings is required depending on the number of metadata schemas involved, and as a consequence problems that appear even in case of one source and one target metadata schema involved, such as different degrees of equivalency (one-to-one, one-to-many, many-to-one and one-to-none), are growing in case of multiple schemas. In addition, Crosswalks proposed, for example from the Library of Congress, are based on the metadata specifications, while real data conversion might be very different and usually requires instructions and explanations to be provided depending on specific documentation expressions [60].

Finally, an additional benefit of using ontologies as an interoperability mechanism is that they can be part of a wider information integration scenario, such as the scenario presented in Chapter 4. What is more, there are other possible approaches, for instance as the Multiple Ontologies Approach [62],[63] where each data source is represented by its own ontology. The implementation of inter-ontology mappings allows the integration of the autonomous data sources originated from related or non related domains.

Concluding, APs reflect the description needs and practices occurred in specific domains. Ontologies aim to define explicitly the nature and the existing norms of a domain as well as to integrate the different description practices occurred inside the domain.

6. Conclusions and further research

Despite the recognition of the importance to use standards for describing PSI and the need to develop PSI systems that implement interoperable documentation methods, the necessary means and guidelines are missing from the PSI community. This paper contributes to the development of PSI integration introducing a semantic interoperability approach. The basic outcomes of this research were to define an Application Profile and an Ontology for documenting PSI and enhancing its ability to interoperate and, moreover, propose the use of the ontology as a mediator in a semantic integration scenario. Based on these efforts we finally discussed the usability of the two metadata approaches in terms of documentation ability and interoperability.

The researchers are further studying in more detail the comparison of various metadata interoperability methods, especially exploring the relationships defined between metadata schemas and ontologies. There are issues to investigate such as (a) the ability of an ontology to encompass the conceptual meanings and complex relationships of a specific domain of interest, and (b) the degree of automation for the creation of semantic mappings between metadata schemas and ontologies. The e-GMS to PSI ontology mapping can be encoded using automated tools, such as OWL editors and XML technologies [43]. Nonetheless, human intervention is required to facilitate describing the semantic mapping, provided that it is a deep conceptual effort.

Our future research work also includes the evaluation of the ontology proposed. The aim of the evaluation process is to validate the completeness of the ontology confirming whether it covers completely the constructs of the domain. At the same time, it would be helpful for the PSI user community to improve the ontology. As part of our evaluation plan we are oriented to follow a data-driven evaluation approach [64] aiming to test the hypothesis of the information loss minimization in the integration process of different metadata schemas due to the exploitation of the semantic relations and similarities between the metadata terms. In particular we aim to try semantic mappings from data sources described by various metadata schemas, such as AGLS and NZGLS, to the ontology. This process will provide useful conclusions about the ontology refinement, its semantic fullness and integration abilities. Besides, we intend to develop an ontology based mediator system that will implement the interoperability scenario presented in this paper. This prototype system will test the quality as well as the performance of the ontology-based metadata integration.

References

- [1] European Commission - Directorate General for the Information Society (2000), Commercial Exploitation of Europe's Public Sector Information, Available at: http://ec.europa.eu/information_society/policy/psi/docs/pdfs/pira_study/commercial_final_report.pdf
- [2] European Commission - Directorate General for the Information Society (2004), Exploiting the potential of Europe's Public Sector Information, Available at: http://ec.europa.eu/information_society/policy/psi/docs/pdfs/brochure/psi_brochure_en.pdf
- [3] Commission of the European Communities (2001), eEurope 2002: Creating an EU Framework for the exploitation of Public Sector Information, Available at: http://mineco.fgov.be/information_society/administrations/eEurope/communication_2001_en.pdf
- [4] G. Prokopiadou, C. Papatheodorou and D. Moschopoulos (2003), "Government Information Centers: Digital Library Architecture for Depicting Public Sector's Hierarchy", in Proceedings of the 3rd European Conference on E-Government, Dublin, Ireland, 4-7 July 2003, pp. 349-358.
- [5] General Secretariat for Information Systems (Ministry of Economy and Finance) (2008), Available at: <http://www.gsis.gr/english/index.html>
- [6] E. Mayo and T. Steinberg (2007), "The Power of Information: An independent review", Available at: <http://www.opsi.gov.uk/advice/poi/power-of-information-review.pdf>
- [7] G. Koutrika (2005), "Heterogeneity in Digital Libraries: two sides of the same coin", Delos Newsletter, No 3, Available at: <http://delos-old.isti.cnr.it/newsletter/issue3/feature2/>
- [8] National Archives of Australia (2002), AGLS Metadata Element Set v 1.3, Available at: <http://www.naa.gov.au/records-management/publications/AGLS-Element.aspx>
- [9] Archives New Zealand (2004), NZGLS Metadata Element Set Version 2.1, Available at: <http://www.e.govt.nz/standards/nzglsl/standard>
- [10] Cabinet Office (2004), e-Government Metadata Standard Version 3.0, Available at: http://www.govtalk.gov.uk/documents/eGovMetadataStandard_20040429.pdf
- [11] DCMI (2006), Dublin Core Metadata Element Set, Version 1.1, Available at: <http://dublincore.org/documents/dces/>
- [12] GILS Maintenance Agency (1997), Global Information Locator Service (GILS), Available at: <http://www.gils.net/>
- [13] R. Heery and M. Patel (2000), "Application profiles: mixing and matching metadata schemas", Ariadne, No 25. Available at: <http://www.ariadne.ac.uk/issue25/app-profiles/>
- [14] M.St. Pierre and W.P.Jr. LaPlant (1998), Issues in Crosswalking Content Metadata Standards, Available at: <http://www.niso.org/press/whitepapers/crswalk.html>
- [15] L. Bountouri and M. Gergatsoulis (2007), "Interoperability between archival and bibliographic metadata", in Proceedings of the First International Workshop on Cultural Heritage on the Semantic Web (in conjunction with the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference), 12 November, Busan, Korea, Available at: <http://www.cs.vu.nl/~laroyo/CH-SW/ISWC-wp9-proceedings.pdf>
- [16] C. Kakali, I. Lourdi, T. Stasinopoulou, L. Bountouri, C. Papatheodorou, M. Doerr and M. Gergatsoulis (2007), "Integrating Dublin Core metadata for cultural heritage collections using ontologies", in Proceedings of the International Conference on Dublin Core and Metadata Applications (DC 2007), Singapore, pp. 128-139.
- [17] T. Stasinopoulou, L. Bountouri, C. Kakali, I. Lourdi, C. Papatheodorou, M. Doerr and M. Gergatsoulis (2007), "Ontology-based Metadata Integration in the Cultural Heritage Domain", in Lecture Notes Computer Science, ICADL 2007, Vol. 4822, Springer, Berlin - Heidelberg, pp. 165-175.

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

- [18] H. Wache, T. Voge, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hubner (2001), "Ontology-based integration of information - a survey of existing approaches", in Proceedings of the Workshop: Ontologies and Information Sharing (IJCAI-01), pp. 108-117, Available at: www.let.uu.nl/~Paola.Monachesi/personal/papers/wache.pdf
- [19] N.F. Noy (2004), "Semantic Integration: a Survey of Ontology-Based Approaches", SIGMOD Record, Vol. 33, No 4, pp. 65-70.
- [20] EAC Working Group (2004), Encoded Archival Context: beta, Available at: <http://www.iath.virginia.edu/eac/>
- [21] Library of Congress (2002), Encoded Archival Description, Available at: <http://www.loc.gov/eac>
- [22] G. Kavadias and E. Tambouris (2003), "GovML: A Markup Language for Describing Public Services and Life Events", in Lecture Notes in Artificial Intelligence, Vol. 2645, Springer, Berlin - Heidelberg, pp. 106-115.
- [23] DCMI (2007), DCMI Abstract Model, Available at: <http://dublincore.org/documents/abstract-model/>
- [24] DCMI (2008), DCMI Metadata Terms, Available at: <http://dublincore.org/documents/dcmi-terms/>
- [25] International Council on Archives (2003), ISAAR (CPF) International Standard Archival Authority Record for Corporate Bodies, Persons and Families, Available at: [http://www.icacds.org.uk/eng/ISAAR\(CPF\)2ed.pdf](http://www.icacds.org.uk/eng/ISAAR(CPF)2ed.pdf)
- [26] DCMI (2000), Using Dublin Core, Available at: <http://dublincore.org/documents/2000/07/16/usageguide/sectc.shtml>
- [27] T. Berners-Lee, J. Hendler, and O. Lassila (2001), "The Semantic Web", Scientific American, Vol. 284 No 5, pp. 34-43.
- [28] G. Antoniou and F. van Harmelen (2003), "Web Ontology Language: OWL", in Handbook on Ontologies in Information Systems, Springer, Berlin - Heidelberg, pp. 76-92.
- [29] I.F. Cruz and H. Xiao (2005), "The Role of Ontologies in Data Integration", Journal of Engineering Intelligent Systems, Vol. 13 No 4, pp. 245-252.
- [30] T. Gruber (1993), "A translation approach to portable ontology specifications", Knowledge Acquisition, Vol. 5 No 2, pp. 199-220.
- [31] C. Partridge (2002), "The Role of Ontology in Integrating Semantically Heterogeneous Databases", Technical report, LADSER-CNR.
- [32] M. Uschold and M. Gruninger (1996), "Ontologies: Principles, Methods and Applications", Knowledge Engineering Review, Vol. 11 No 2, pp. 93-155.
- [33] P. Lehti and P. Fankhauser (2004), "XML Data Integration with OWL: experiences and challenges", Proceedings of SAINT 2004, pp. 160-170.
- [34] International Council on Archives (2006), ISAF: International Standard for Activities/Functions of Corporate bodies, Available at: http://www.ica.org/sites/default/files/ISAF_ENG.pdf
- [35] International Council on Archives (2000). ISAD(G): General International Standard Archival Description, Available at: http://www.ica.org/sites/default/files/isad_g_2e.pdf
- [36] L. Duranti (1998), Keeping the records straight, Available at: <http://www.publicaffairs.ubc.ca/ubcreports/1998/98mar05/98mar5pro.html>
- [37] IEEE (2004), Suggested Upper Merged Ontology (SUMO), Available at: <http://www.ontologyportal.org/>
- [38] ICOM/CIDOC (2007), Definition of the CIDOC Conceptual Reference Model, Available at: http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.2.doc
- [39] Cognitive Science Laboratory Princeton University (2006), WordNet, Available at: <http://wordnet.princeton.edu/>

- [40] U.S. Census Bureau (2002), North American Industry Classification System (NAICS), Available at: <http://www.census.gov/epcd/naics02/>
- [41] State Services Commission of New Zealand (2004), Functions of New Zealand (FONZ) Thesaurus v 2.30, Available at: <http://www.e.govt.nz/standards/nzxls/thesauri/fonz/fonz-2-30.pdf>
- [42] National Archives of Australia (2005), Australian Governments' Interactive Functions Thesaurus, Available at: <http://www.naa.gov.au/records-management/create-capture-describe/describe/AGIFT/index.aspx>
- [43] H. Kondylakis, M. Doerr and D. Plexousakis (2006), "Mapping Language for Information Integration, Technical Report 385", ICS-FORTH, Available at: http://www.ics.forth.gr/ftp/tech-reports/2006/2006.TR385_Mapping_Language_Information_Integration.pdf
- [44] European Committee for Standardization (2003), Dublin Core eGovernment Application Profiles, Available at: <ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14860-00-2003-Nov.pdf>
- [45] MIREG and DCMI Government Working Group (2001), Government Application Profile, Available at: <http://dublincore.org/documents/2001/09/17/gov-application-profile/>
- [46] S.-K. Han, H.-S. Lee and Y.-S. Jeong (2006), "Conceptual Model of Metadata Schema for Records Management", in Proceedings of the 2nd International Symposium on Knowledge processing and Service for China, Japan and Korea, Metadata And Ontology, Beijing, China, pp. 21-31.
- [47] ISO (2001), ISO 15489-1:2001, Information and documentation -- Records management -- Part 1: General.
- [48] ISO (2001), ISO/TR 15489-2:2001, Information and documentation -- Records management -- Part 2: Guidelines.
- [49] ISO (2006), ISO 23081-1:2006, Information and documentation -- Records management processes -- Metadata for records -- Part 1: Principles.
- [50] B. Bercic and M. Vintar (2003), "Ontologies, Web Services, and Intelligent Agents: Ideas for Further Development of Life-Event Portals", in Lecture Notes in Computer Science, Vol. 2739, Springer, Berlin - Heidelberg, pp. 329-334.
- [51] L. Stojanovic, G. Kavadias, D. Apostolou, F. Probst and K. Hinkelmann (2004), Ontology-enabled e-Gov Service Configuration, Available at: http://www.hsw.fhso.ch/ontogov/documents/D2_EGov_Service_Lifecycle_Ontology.pdf
- [52] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari and L. Schneider (2003), "WonderWeb Deliverable D17: The WonderWeb Library of Foundational Ontologies", Preliminary Report, Available at: <http://wonderweb.semanticweb.org/deliverables/documents/D17.pdf>
- [53] DCMI (2008), DCMI Description Set Profile, Available at: <http://dublincore.org/architecture/wiki/DescriptionSetProfile>
- [54] M. Doerr, C. Lagoze, J. Hunter, and T. Baker (2002), "Building Core Ontologies: a White Paper of the DELOS Working Group on Ontology Harmonization", Technical report, DELOS Network of Excellence on Digital Libraries.
- [55] C. Lagoze and J. Hunter (2002), "The ABC Ontology and Model", Journal of Digital Information, Vol. 2 No 2, Article No. 77.
- [56] C. Lagoze, J. Hunter and D. Brickley (2000), "An Event-Aware Model for Metadata Interoperability", in Proceedings of the ABC Workshop, Available at: <http://www.cs.cornell.edu/lagoze/papers/ev.pdf>
- [57] A. Maier, J. Aguado, A. Bernaras, I. Laresgoiti, C. Pedinaci, N. Peña and T. Smithers (2003), "Integration with Ontologies", in Proceedings of the WM2003, Luzern.
- [58] S. Suwanmanee, D. Benslimane and P. Thiran (2005), "OWL-Based approach for Semantic Interoperability", in Proceedings of the AINA, pp. 145-150.

Lina Bountouri, Christos Papatheodorou, Vasilis Soulikias, Mathios Stratis

- [59] R. Tous and J. Delgado (2005), "RDF Databases for Querying XML. A Model-mapping Approach", in Proceedings of the DISWeb 2005 International Workshop Data Integration and the Semantic Web (DISWeb 2005) Portugal.
- [60] L.M. Chan and M.L. Zeng (2006), "Metadata Interoperability and Standardization – A Study of Methodology Part I Achieving Interoperability at the Schema Level", D-Lib Magazine, Vol. 12 No 6, Available at: <http://www.dlib.org/dlib/june06/chan/06chan.html>
- [61] E. Duval, W. Hodgins, S. Sutton and S.L. Weibel (2002), "Metadata Principles and Practicalities", D-Lib Magazine, Vol. 8 No 4, Available at: <http://www.dlib.org/dlib/april02/weibel/04weibel.html>
- [62] N. Noy and H. Stuckenschmidt (2005), "Ontology Alignment: An annotated Bibliography", in Dagstuhl Seminar Proceedings 04391, Semantic Interoperability and Integration, Available at: <http://drops.dagstuhl.de/opus/volltexte/2005/48/pdf/04391.StuckenschmidtHeiner1.Paper.48.pdf>
- [63] D. Calvanese, G. De Giacomo and M. Lenzerini (2001), "A Framework for Ontology Integration", in Proceedings of the First Semantic Web Working Symposium", pp. 303-316.
- [64] J. Brank, M. Grobelnik and D. Mladenić (2005), "A Survey of Ontology Evaluation Techniques", in Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD2005) Slovenia, Available at: <http://kt.ijs.si/dunja/sikdd2005/Papers/BrankEvaluationSiKDD2005.pdf>